Lending Club Case Study

Upgrad
Prepared by:
     Priyanuj Misra
     Pratiksha Aggarwal

Content

Identification of risky loan applicants - Lending Club Analysis

GAP
- The goal of the case study is to identify key factors that indicate loan defaults using exploratory data analysis (EDA) to reduce credit loss and improve risk assessment.

Current State
- LC is the largest online loan marketplace, offering personal, business, and medical procedure loans through an easy-to-use platform.
- A major financial issue for the company is credit loss by risky borrowers who default on loans
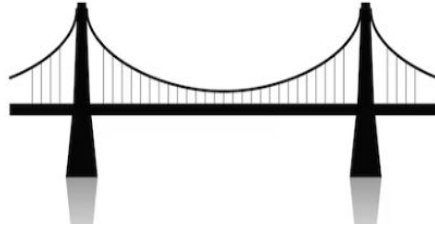
Desired Future State
- LC is able to identify high risk customers via automated capability system
- LC loan processors have additional information about risky customers before processing the loan

Key Question
- What are the key factors in defining customer loan risk?
- Segments more prone to loan defaults

Importance of Risk Analytics

01

Minimizing Financial Risk

- Risk analytics helps fintech companies identify high-risk borrowers, preventing defaults and minimizing credit loss

02

Enhancing Customer Segmentation and Personalized Offers

- Enables personalized loan offers based on risk profiles and ensures regulatory compliance for sustainable business operations.

03

Framework for future risk management

- Risk analytics helps fintechs comply with regulations, avoid penalties, manage risk, and optimize capital allocation for a sustainable business

# Analytical Steps

Analytical Processes Involved

Data Understanding

1. Preliminary EDA
2. In-depth EDA
3. Outlier Detection
4. Univariate Analysis
5. Bivariate Analysis
6. Data Imputation
7. Data Manipulation
8. Data Extraction
9. Findings and Insights

Findings and Insights

Loan applicants with lower annual income brackets are at higher risk of default

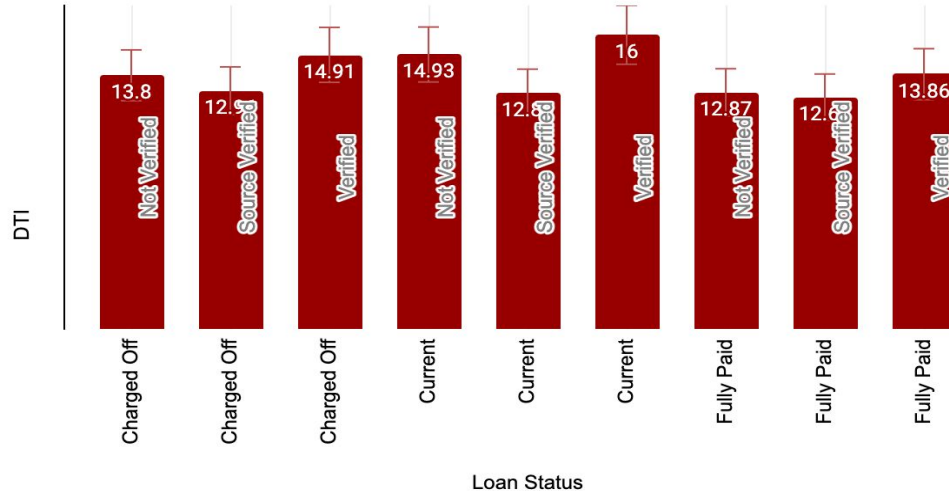Variation of Annual Income per Loan Status Category
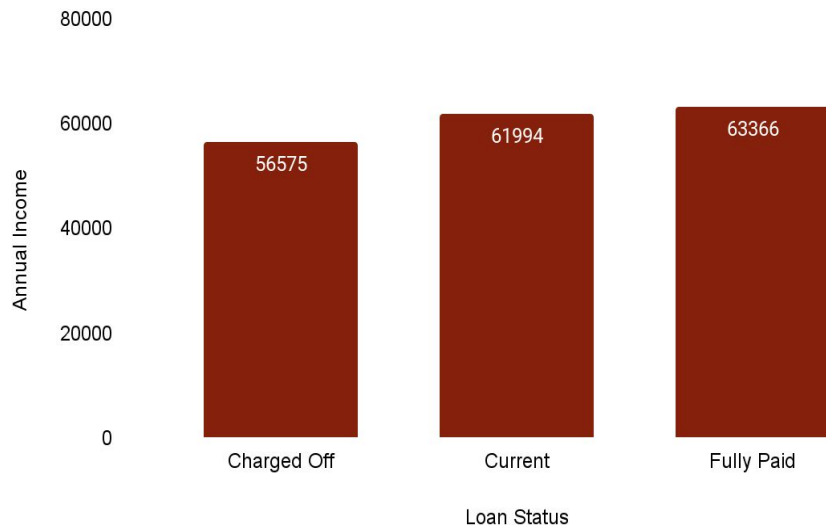
Findings:

1. On an average, customers with annual income of ~ $ 550 K are at a higher risk of default
2. A higher annual income of $ 600 K or above are usually at a lower risk of default

Insights:

1. Customers with lower income might find it difficult to repay off a loan depending on various other factors such as DTI etc

A higher Debt-to-Income ratio (DTI) may be considered as an important factor for risk assessment

DTI across Loan Status and Verification Status

DTI

| Charged Off | Charged Off | Charged Off | Current | Current | Current | Fully Paid | Fully Paid | Fully Paid |
| Not Verified 13.8 | Source Verified 12.9 | Verified 14.91 | Not Verified 14.93 | Source Verified 12.8 | Verified 16 | Not Verified 12.87 | Source Verified 12.6 | Verified 13.86 |

Loan Status

Findings:
1. On an average, customers with annual income of ~ $ 550 K are at a higher risk of default
2. A higher annual income of $ 600 K or above are usually at a lower risk of default

Insights:
1. Customers with lower income might find it difficult to repay off a loan depending on various other factors such as DTI etc

Note: Higher DTI percentages for Current loan status category indicates that there are customers who will have a higher tendency to default in the future. Hence comparison is between Fully Paid and Charged Off only

Customers with higher average installment (~ 300) are more likely to default

Loan Status	Average Installment
Charged Off	~ $ 300
Current	~ $ 269
Fully Paid	~ $ 292

Drill down analysis post segmentation of customers into deciles based on last year sales

1. Observed an average discount percentage of approximately $ 300 for the charged off segment
2. We can have a capping based on business discussion to check if we can put a threshold for EMI and Loan Disbursement amount based on the annual income of the customers
3. There Current Segment needs more understanding of the loan disbursement process

For States IN, ME and IA, all customers have fully paid off their loan
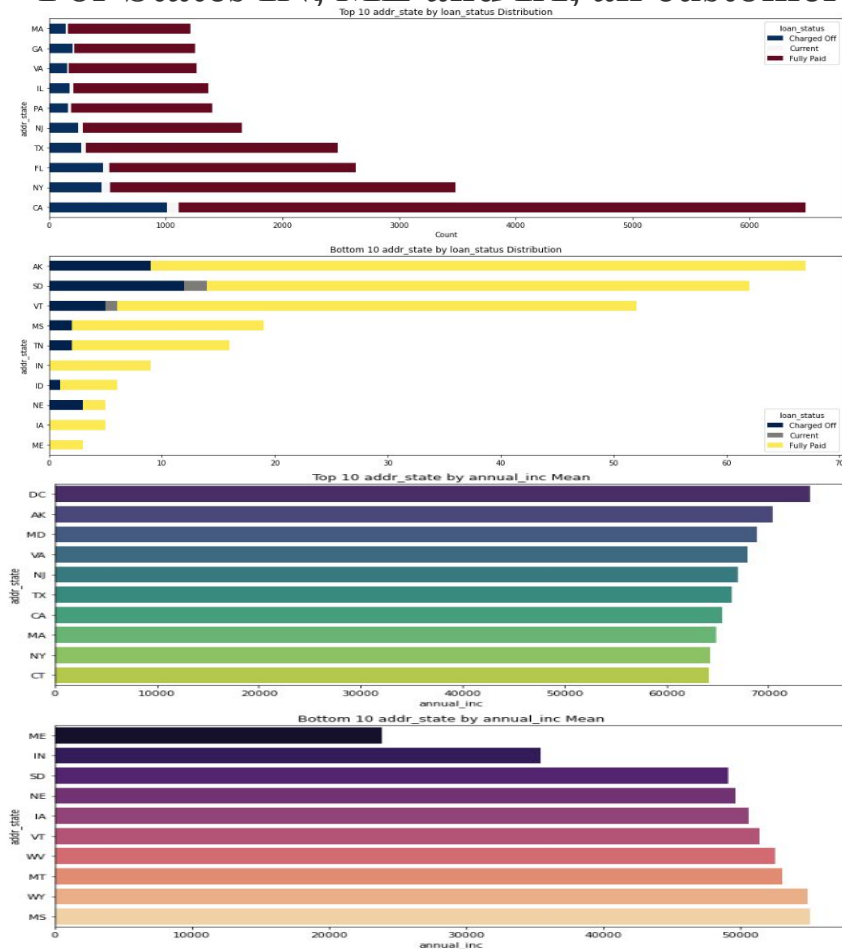
Observations

1. Plotting top 10 and bottom 10 states by loan status, we observed states IN, ME and IA have no defaulters
2. For the same comparing Annual Income by state shows that ME, IN and IA are low income households as well
3. This can be associated with the overall literacy, and other factors based on the area
4. We can definitely improvise the thresholds per state
5. CA is among the top 10 highest annual income, greater than $ 60K, has highest fully paid customers as well. However, due to high income certain customers might take a loan exceeding their capability as charged off is higher as well

Risk Strategies and Recommendations

Summary of Analytics and Future course of action

Summary

1. A threshold/methodology needs to be incorporated to balance annual income, amount given for loan, DTI and EMI
2. Analyze Fully Paid vs Charged Off wherever necessary and leave the category Current
3. Interest rates are usually less than 12% for Fully Charged whereas for defaulters it is close to 13-14% as mean.
4. Revolving Utility is higher for Charged Off Category (>50) while for Fully paid it is less than 50

Future Course of Action

1. To avoid manual dependency for later purpose, identify more important features, use predictive analytics to classify customers as Charged Off or Not.
2. Build a dashboard for inhouse loan processors, once they fill in relevant information, the dashboard should output the maximum fund disbursement amount and EMI value for customers

Recommendations

1. Prepare a rule based approach based on Annual Income, DTI, Revolving Balance etc to gauge the risk for the customers accessible by the loan processors
2. Provide a drop down for Purpose Column and provide only relevant categories to choose from. Due to randomness in that column further analysis was not possible to understand loan default based on wants vs needs or other purposes
3. Delinquency details should be recorded properly as count of delinquency per customer etc such that this can be taken care of in further loan processing
4. New vs Returning customer segmentation will help to understand which customers will bring in more money and have a better record of paying off their debts well
5. Areas with higher Fully Paid off customers can apply higher interest rates for high risk customers because based on history most of the customers will pay it off
6. Additionally, for the long term, predictive analytics to segment customers, understand their inhouse credit score etc can be planned to streamline the process flow

Thank You